

Machine Learning based estimation of Chlorophyll and Flavonoid content in Bitter Leaf using Color and GLCM Texture Features

Retno Damayanti^{1,2,*}, Yusuf Hendrawan², Sandra² and Bambang Dwi Argo²

¹Department of Agroindustrial Technology, Universitas Brawijaya, Malang 65145, Indonesia; ²Department of Biosystems Engineering, Universitas Brawijaya, Malang 65145, Indonesia

*Corresponding author's e-mail: damayanti@ub.ac.id

The prediction of plant biochemical content is critical for applications in both agriculture and pharmacology. This study presented a non-invasive approach that integrates Gray Level Co-occurrence Matrix (GLCM) texture analysis with Artificial Neural Networks (ANNs) to estimate chlorophyll and flavonoid content in *Vernonia amygdalina* (bitter leaf). Leaves at three developmental stages were analyzed using standard spectrophotometric methods for validation. GLCM features (contrast, correlation, energy, and homogeneity) were extracted from normalized leaf images, and feature selection techniques were employed to identify the most predictive variables. An optimized ANN, trained and validated using these selected features, demonstrated high predictive accuracy without relying on destructive sampling methods. This integrated GLCM-ANN framework offers improved efficiency, reduced costs, and enhanced scalability compared to conventional approaches. Its potential applications extend to precision agriculture enabling real-time monitoring of plant health, early stress detection, and optimized nutrient management as well as to the pharmaceutical industry for improving the quality control of herbal medicines. Future work will focus on expanding the dataset and incorporating multi-modal imaging to further refine model performance and advance non-invasive plant biochemical analysis.

Keywords: Bitter leaf, photosynthesis, non-destructive biochemical analysis, hyperspectral imaging, regression models, Artificial Neural Networks, precision agriculture.

INTRODUCTION

The accurate prediction of plant chlorophyll and flavonoid content plays a pivotal role in agriculture, pharmacology, and environmental studies. Chlorophyll, essential for photosynthesis, and flavonoids, key antioxidants, are particularly relevant in bitter leaf (*Vernonia amygdalina*) due to its rich bioactive profile and importance in traditional medicine. The development of efficient, non-invasive methods to quantify these compounds has therefore become a priority, as bitter leaf's phytochemical properties influence its therapeutic potential and agricultural value (Rusadi, 2023; Rivero-Montejo *et al.*, 2021).

Current techniques for non-destructive biochemical analysis include spectrophotometry and hyperspectral imaging. While reliable, these methods can be time-intensive, sensitive to environmental variability, and require complex data interpretation (Yan *et al.*, 2021). Hyperspectral imaging captures data across a broad spectrum, enabling detailed

spatial and spectral analysis of plant tissues. However, these methods face significant limitations. Spectrophotometry, while precise, is often time-intensive and constrained by environmental variability and the need for laboratory conditions. Hyperspectral imaging, despite its potential to map biochemical distribution, struggles with interpreting data from heterogeneous plant tissues due to uneven compound distribution and external interferences, such as ambient light and temperature fluctuations (Yan *et al.*, 2021).

To overcome these challenges, machine learning (ML) has emerged as a powerful tool, enabling the integration of image-derived features and complex datasets to improve predictive accuracy (Benos *et al.*, 2021; Adebiyi *et al.*, 2020). ML applications in plant studies have successfully optimized classification and regression models for traits like chlorophyll content, drought tolerance, and crop yield (Victoire, 2023). Yet, ML models often demand extensive training data and may struggle to generalize across diverse conditions (Nabwire *et al.*, 2021).

Damayanti, R., Y. Hendrawan, Sandra and B.D. Argo. 2025. Machine Learning based estimation of Chlorophyll and Flavonoid content in Bitter Leaf using Color and GLCM Texture Features. Journal of Global Innovations in Agricultural Sciences 13:1011-1019.

[Received 14 Sep 2024; Accepted 15 Jan 2025; Published 21 Jun 2025]



Attribution 4.0 International (CC BY 4.0)

Table 1. GLCM features used for texture analysis.

GLCM Feature	Mathematical Definition	Biological Relevance
Contrast	$\sum_{i,j} (i - j)^2 P[i, j]$	Measures the local variations in the gray-level co-occurrence matrix. High contrast indicates significant intensity variations, correlating with higher flavonoid content due to increased structural complexity in leaves.
Correlation	$\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)P[i, j]}{\sigma_i \sigma_j}$	Reflects the linear dependency of gray levels on those of neighboring pixels. Higher correlation may be associated with specific biochemical compositions in the leaves.
Energy	$\sum_{i,j} P[i, j]^2$	Represents the uniformity of the texture. High energy values indicate homogeneous textures, which may correlate with higher chlorophyll content due to uniform chloroplast distribution.
Homogeneity	$\sum_{i,j} \frac{P[i, j]}{1 + i - j ^2}$	Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. High homogeneity indicates smoother textures, possibly related to higher chlorophyll levels.
Entropy	$-\sum_{i,j} P[i, j] \log P[i, j]$	Measures the randomness in the texture. High entropy indicates complex textures, which may be linked to increased flavonoid accumulation as a response to stress factors.
Inverse Difference Moment (IDM)	$\sum_{i,j} \frac{P(i, j)}{1 + (i, j)^2}$	Similar to homogeneity, emphasizes contributions from neighboring pixels with similar intensity values, indicating texture smoothness. Higher IDM may correlate with higher chlorophyll content.

Note: In the mathematical definitions, $P(i, j)$ represents the normalized probability values in the GLCM, i and j are the pixel intensity values, μ_i and μ_j are the means, and σ_i and σ_j are the standard deviations of the row and column sums of the GLCM.

Integrating Gray Level Co-occurrence Matrix (GLCM) texture analysis with Artificial Neural Networks (ANNs) offers a promising solution to these issues. GLCM quantifies spatial pixel relationships in leaf images, capturing subtle texture patterns linked to biochemical properties (Baker and Sardari, 2021; Kumar *et al.*, 2022). However, the computational demands of GLCM and the need for advanced image processing capabilities can pose challenges, particularly in field applications where resource constraints are common (Zivcak *et al.*, 2017). By leveraging GLCM-derived features, ANNs can model non-linear relationships between leaf texture and compound levels, reducing reliance on intensive calibration and enhancing robustness in varying environments. Bitter leaf's distinctive phytochemical composition makes it an ideal model for validating this approach, as understanding chlorophyll and flavonoid patterns can inform cultivation practices and improve product consistency in both agriculture and pharmacology (Paponov *et al.*, 2021). Integrating texture analysis with ML models such as ANNs allows for the simultaneous analysis of multiple image features, improving the precision and scalability of predictions. ANNs, inspired by the neural architecture of the human brain, excel in modeling complex relationships between input features and output variables, making them ideal for plant biochemical studies (Benos *et al.*, 2021).

This study addresses the limitations of current techniques by employing a GLCM-ANN framework to estimate chlorophyll and flavonoid content in bitter leaf, providing a non-invasive, scalable, and cost-effective analytical strategy. The findings

are expected to broaden our understanding of leaf texture-biochemistry relationships, thus advancing precision agriculture and supporting the sustainable development of plant-based medicinal resources.

MATERIALS AND METHODS

Study design and plant sample preparation: Bitter leaf (*Vernonia amygdalina*) samples were collected from multiple locations in Malang, Indonesia. The selection of plant samples focused on the leaves, given their high concentrations of chlorophyll and flavonoids (Sahin *et al.*, 2022). Samples were harvested from three different maturity levels: top (young leaves), middle (mature leaves), and bottom (old leaves). All samples were collected following institutional guidelines and relevant regulations, ensuring ethical and sustainable sourcing practices. To ensure consistency, all samples were collected at the same time of day (during the early morning) under similar environmental conditions to minimize diurnal variations in biochemical content (Kim, 2020). After collecting, leaves were washed with distilled water to remove surface contaminants, air-dried at room temperature, and stored in labeled plastic bags. Care was taken to handle samples with clean tools to prevent contamination and degradation (Derbalah, 2012).

Image acquisition: Digital images of bitter leaves were captured using a Canon EOS 700D camera mounted inside a controlled black box environment (Fig. 1). The black box minimized external light interference, providing consistent imaging conditions. The camera settings were standardized at



an aperture of f/5.6, an exposure time of 1/3 second, and ISO 100. Leaf samples were placed against a black fabric background for contrast enhancement. Captured images were cropped, normalized to a consistent scale and intensity range, and then processed to extract color features in RGB, HSV, HSL, and Lab* color spaces. GLCM texture features such as contrast, correlation, energy, and homogeneity were computed from grayscale-converted images using MATLAB R2021a, leveraging established equations (Butnariu and Sarac, 2018; Annadurai and Nelson, 2018). Feature selection algorithms (Chi-Square and ReliefF in WEKA 3.8) were applied to identify the most predictive texture variables for downstream modeling.

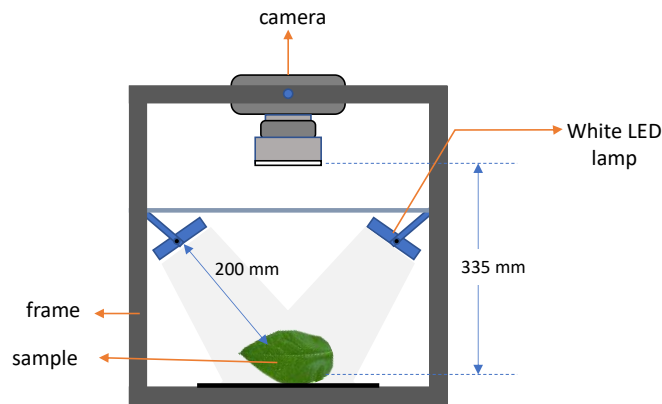


Figure 1. Black box setup for image acquisition.

Biochemical content analysis

Chlorophyll analysis: Total chlorophyll content was quantified using Arnon's method with slight modifications (Arnon, 1949). Two grams of leaf tissue were ground in a mortar with 10 mL of acetone (PA grade) and left to extract for 24 hours at 4°C. The extract was filtered (Whatman No. 42), and absorbance was measured at 645 nm and 663 nm using a Shimadzu UV-Vis spectrophotometer. Chlorophyll a, Chlorophyll b, and Total Chlorophyll concentrations were calculated using standard equations.

$$\text{Chlorophyll a} = 12.21 \times A_{663} - 2.81 \times A_{645}$$

$$\text{Chlorophyll b} = 20.13 \times A_{645} - 5.03 \times A_{663}$$

$$\text{Total Chlorophyll} = \text{Chlorophyll a} + \text{Chlorophyll b}$$

Flavonoid analysis: Flavonoid content was determined using the aluminum chloride colorimetric method (Mahmoud, 2023). One gram of leaf samples was macerated in 10 mL of 80% ethanol for 4 hours, centrifuged at 5000 rpm for 20 minutes, and the supernatant reacted with 0.5 mL of 10% AlCl_3 , 0.5 mL of 1M NaOH, and 2 mL of distilled water. After incubating for 30 minutes at room temperature, absorbance was measured at 415 nm, and flavonoid concentration was expressed as mg QE/g.

Feature selection: Texture features derived from GLCM were evaluated for relevance to chlorophyll and flavonoid

content prediction. Four key GLCM features; contrast, correlation, energy, and homogeneity, were prioritized based on their ability to capture texture variation (Butnariu and Sarac, 2018; Sinha, 2014). The mathematical definitions and biological relevance of each GLCM feature can be seen in Table 1. Feature selection was performed using the Chi-Square and ReliefF algorithms implemented in the Waikato Environment for Knowledge Analysis (WEKA 3.8). The selected features were ranked by their contribution to model accuracy, and the top-performing features were used as input for ANN modeling.

ANN model development and optimization: The dataset, comprising 1010 samples, was divided into training (70%) and validation (30%) sets. An ANN model was constructed using selected GLCM features as inputs, optimized through the Levenberg-Marquardt training algorithm (Annadurai and Nelson, 2018). Different network architectures (1–2 hidden layers, 10–40 nodes per layer) were tested. Mean Square Error (MSE) and the coefficient of determination (R^2) were used to assess performance. To ensure reproducibility, code and additional data analysis scripts are available upon reasonable request.

RESULTS

Biochemical Content Across Developmental Stages

Chlorophyll content: The analysis revealed distinct variation in chlorophyll content across the developmental stages of *Vernonia amygdalina* leaves (Table 2). Chlorophyll levels were highest in the older leaves (0.231 mg/g) compared to the mature (0.214 mg/g) and younger leaves (0.138 mg/g). This pattern suggests that, in *V. amygdalina*, chlorophyll accumulation persists longer in older leaves, potentially reflecting species-specific leaf physiology. Previous findings have indicated that chlorophyll typically peaks during full leaf development and decreases during senescence (Sheng et al., 2021). However, our results imply a delayed degradation process in this species, particularly in older leaves, which may be influenced by nutrient allocation or leaf age structure within the canopy.

Table 2. Chlorophyll and flavonoid content across developmental stages.

Leaf Stage	Chlorophyll (mg/g)	Flavonoids (mg QE/g)
Young (Upper)	0.138 ± 0.02	07.655 ± 0.15
Mature (Middle)	0.214 ± 0.03	10.728 ± 0.20
Old (Lower)	0.231 ± 0.02	11.743 ± 0.18

Flavonoid content: Flavonoid concentrations also varied significantly across leaf maturity stages (Table 2). Older leaves showed the highest flavonoid levels (11.7433 mg QE/g), followed by mature leaves (10.7282 mg QE/g), and young leaves (7.655 mg QE/g). This trend aligns with the



protective role flavonoids play in response to environmental stresses, which may accumulate more prominently in older leaves. The intermediate concentrations in mature leaves indicate a transitional phase where both biosynthesis and metabolic conversion contribute to overall flavonoid balance. These findings underscore the importance of developmental stage in determining the phytochemical composition of bitter leaf.

Correlation between biochemical content and texture features: Texture analysis using GLCM features revealed significant correlations with both chlorophyll and flavonoid levels (Table 3). Energy and homogeneity were strongly positively correlated with chlorophyll ($R > 0.85$, $p < 0.01$), suggesting that smoother, more uniform textures are associated with enhanced chlorophyll content. Conversely, contrast and correlation showed strong positive associations with flavonoid levels ($R > 0.78$, $p < 0.01$), implying that greater structural complexity may reflect stress-related flavonoid accumulation. These results support the hypothesis that GLCM-derived texture metrics can serve as reliable indicators of underlying biochemical variations in bitter leaf tissues.

Table 3. Correlation between texture features and biochemical content.

Feature	Chlorophyll Correlation (R)	Flavonoid Correlation (R)
Contrast	0.62	0.78
Correlation	0.71	0.81
Energy	0.85	0.68
Homogeneity	0.87	0.63

ANN model performance: From the Lab color space, 10 selected features (e.g., Lab_b Entropy, Lab_L Entropy, Lab_b Energy) emerged as the most predictive inputs for ANN modeling. Feature selection using Chi-Squared attribute evaluation confirmed their importance, with the highest-ranking features demonstrating the strongest correlations to biochemical content (Table 4). When used as inputs to the optimized ANN model, these features yielded high predictive accuracy, eliminating the need for destructive sampling methods. The best ANN architecture (two hidden layers with 30 nodes each) achieved a training MSE of approximately 0.0100 and a validation MSE of about 0.0839, indicating robust model performance and generalizability.

Table 4. Feature selection results.

Image features	Weight	Rank
Lab_b Entropy	2193.24	1
Lab_L Entropy	1836.27	2
Lab_b Energy	1813.31	3
Lab_b Homogeneity	1714.21	4
Lab_b Inverse	1690.62	5
Lab_L Homogeneity	1543.52	6
Lab_L Inverse	1492.57	7
Lab_a Homogeneity	1419.27	8
Lab_a Inverse	1412.25	9
Lab_a Entropy	1381.67	10

Table 5. ANN modeling results based on image features ranking.

Input	MSE training	MSE validation
Rank 1 - 2		
Rank 1 - 3		
Rank 1 - 4	0.0100	0.1025

Table 6. ANN training and validation results with various learning functions.

No.	Learning function	R Training	R Validation	MSE Training	MSE Validation
1	Traincgb (Conjugate Gradient Backpropagation with Powell-Beale Restarts)	0.99254	0.98855	0.0100	0.1174
2	Traincgf (Conjugate Gradient Backpropagation with Fletcher-Reeves Updates)	0.98617	0.98556	0.0189	0.1526
3	Traincgp (Conjugate Gradient Backpropagation with Polak Ribiere Update)	0.99270	0.99287	0.0100	0.1071
4	Traingd (Gradient Descent Backpropagation)	0.86166	0.98285	0.1649	0.1837
5	Traingda (Gradient Descent with Adaptive Learning Rate Backpropagation)	0.87915	0.98535	0.1444	0.1539
6	Traingdm (Gradient Descent with Momentum Backpropagation)	0.86124	0.98344	0.1667	0.2017
7	Traingdx (Gradient Descent with Momentum and Adaptive Learning Rate Backpropagation)	0.87210	0.98019	0.2007	0.2158
8	Trainlm (Lavenberg-Marquadt Backpropagation)	0.99265	0.99357	0.0100	0.0839
9	Trainoss (One-Step Secant Backpropagation)	0.98661	0.99170	0.0182	0.1092
10	Trainrp (Resilient Backpropagation)	0.98503	0.99257	0.0195	0.0912
11	Trainscg (Scaled Conjugate Gradient Backpropagation)	0.99278	0.98822	0.0100	0.1351



Table 7. Results of ANN training and validation with various activation functions.

Learning function	Fungsi Aktivasi			R		MSE	MSE
	Hidden Layer 1	Hidden Layer 2	Output Layer	Training	Validation	Training	Validation
Trainlm	Tansig	Tansig	Purelin	0.9922	0.98631	0.0099	0.1610
	Tansig	Tansig	Tansig	0.99265	0.99357	0.0100	0.0839
	Tansig	Tansig	Logsig	0.94616	0.96931	0.3110	0.4203
	Logsig	Logsig	Purelin	0.9923	0.99055	0.0098	0.1249
	Logsig	Logsig	Tansig	0.99267	0.98948	0.0099	0.1160
	Logsig	Logsig	Logsig	0.94803	0.96901	0.3099	0.4420

Table 8. ANN training and validation results with various ANN structures.

ANN Structure	R Training	R Validation	MSE Training	MSE Validation
10-30-2	0.99332	0.97280	0.0100	0.2943
10-40-2	0.99319	0.97872	0.0100	0.2261
10-30-30-2	0.99265	0.99357	0.0100	0.0839
10-30-40-2	0.99256	0.99424	0.0100	0.1101
10-40-40-2	0.99273	0.99371	0.0096	0.1160

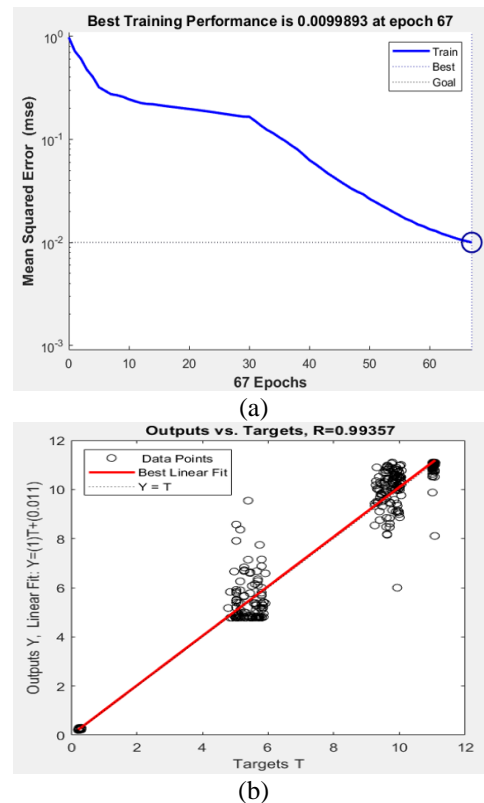
Rank 1 - 5	0.0100	0.7490
Rank 1 - 6	0.0100	0.6320
Rank 1 - 7	0.0099	0.5190
Rank 1 - 8	0.0098	0.3228
Rank 1 - 9	0.0099	0.3418
Rank 1 - 10	0.0100	0.0839

Sensitivity analyses confirmed that the Levenberg-Marquardt algorithm, combined with ‘tansig’ activation functions, produced the most reliable results (Table 6, 7, and 8). Training converged within a relatively small number of epochs, and regression analyses showed strong agreement between predicted and measured values (R training > 0.99, R validation > 0.99). These findings highlight the potential of integrating GLCM texture features and ANN modeling for efficient, non-destructive estimation of chlorophyll and flavonoid content in bitter leaf, with implications for real-time monitoring and improved quality control in both agricultural and pharmacological applications.

The learning curves (Fig. 2a) showed rapid convergence during training, with MSE values stabilizing after approximately 67 epochs. Validation results indicated strong predictive accuracy, with R -values of 0.99357 for chlorophyll and flavonoids (Fig. 2b), suggesting a robust fit between the model and the data.

The ANN architecture of chlorophyll and flavonoid showed in Fig. 3, where it consists of 10 input neurons, 2 hidden layers, and 2 output neurons. According to research conducted by [Ogunbo et al. \(2020\)](#), increasing the number of layers in the hidden layer could produce better performance because the number of weights and biases can support the training process of artificial neural networks. The use of multiple hidden layers in the same iteration could speed up

the training process compared to the training process with a single hidden layer ([Zhang and Shen, 2019](#)).

**Figure 2. Training results; (a) epoch and MSE relationship, (b) regression validation plot.**

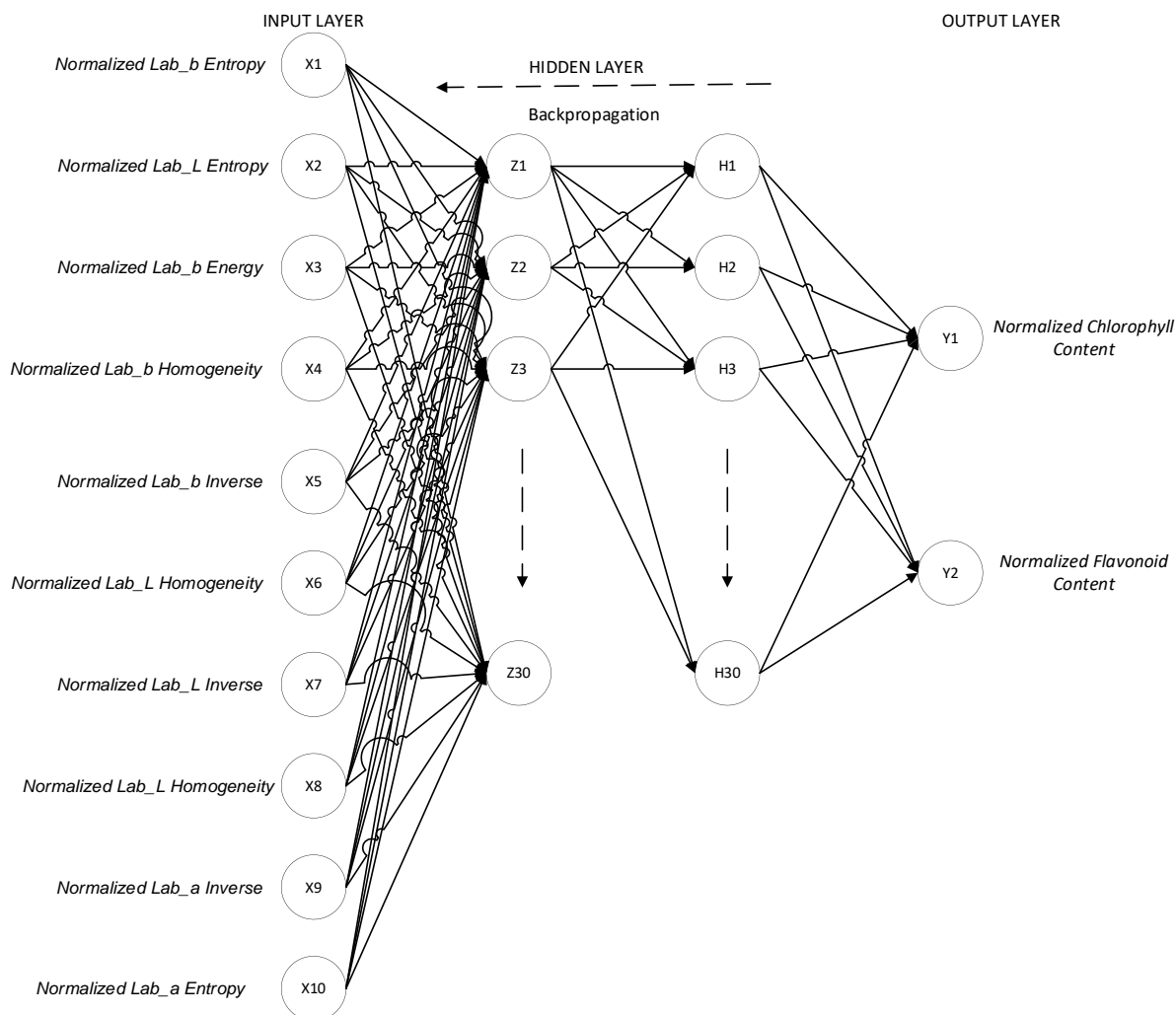


Figure 3. ANN architecture of correlation between texture features and biochemical content.

DISCUSSION

This study demonstrates the notable potential of integrating Gray Level Co-occurrence Matrix (GLCM) texture features with Artificial Neural Network (ANN) models to predict chlorophyll and flavonoid content in *Vernonia amygdalina*. By linking non-invasive texture analysis with advanced computational modeling, this approach provides a more efficient, accurate, and scalable alternative to traditional methods. The findings underscore the critical role of texture features in enhancing model performance and offer valuable insights into the relationship between leaf texture and biochemical properties.

The results revealed distinct patterns in chlorophyll and flavonoid content across leaf developmental stages. Chlorophyll content peaked in older leaves (0.231 mg/g) and was lower in mature (0.214 mg/g) and younger leaves (0.138 mg/g). This species-specific pattern suggests a delayed

chlorophyll degradation process in older *V. amygdalina* leaves, potentially linked to unique physiological traits. This contrasts with prior studies suggesting that chlorophyll accumulation intensifies during leaf maturation and diminishes due to degradation during senescence (Sheng *et al.*, 2021). Flavonoid content displayed an inverse trend, highest in older leaves (11.7433 mg QE/g). This supports the hypothesis that flavonoids serve as protective agents against oxidative stress and environmental challenges, which accumulate with age (He *et al.*, 2021; Guo *et al.*, 2020). The intermediate levels in mature leaves (10.7282 mg QE/g) highlight a balance between biosynthesis and metabolic conversion. Such findings emphasize how leaf age and structural complexity collectively shape phytochemical profiles.

Incorporating GLCM features—contrast, correlation, energy, and homogeneity—into ANN models elucidated the intricate link between texture and biochemical content. High energy



and homogeneity aligned with elevated chlorophyll levels, indicating that smoother leaf surfaces may optimize light capture and chloroplast distribution (Chang *et al.*, 2018). Conversely, higher contrast and correlation corresponded with increased flavonoids, suggesting that more complex leaf textures may foster enhanced protective responses against environmental stressors (Zivcak *et al.*, 2017). These correlations validate the concept that texture attributes derived from GLCM can serve as reliable proxies for underlying biochemical characteristics (Zhang *et al.*, 2022; Roitsch *et al.*, 2019).

The optimal ANN architecture, featuring two hidden layers with 30 neurons each, efficiently modeled the non-linear relationships between GLCM features and biochemical content (Li *et al.*, 2010). This synergy of GLCM-ANN integration streamlines the analytical process, providing rapid and cost-effective predictions without destructive sampling. Such advantages are highly relevant for precision agriculture, where accurate, real-time assessments of chlorophyll and flavonoid content can inform dynamic resource management, crop quality optimization, and improved yield strategies (Sampaio *et al.*, 2021).

These findings align with previous studies using GLCM features in plant analysis (Zhang *et al.*, 2022; Roitsch *et al.*, 2019) and extend the application to a medicinally significant species. By focusing on *Vernonia amygdalina*, this work not only confirms the feasibility of the GLCM-ANN framework but also supports its broader utility for standardizing phytochemical content in commercially and pharmacologically important plants. The observed texture-biochemistry relationships confirm that leaf morphological traits influence functional properties, such as photosynthetic capacity and antioxidant defense mechanisms. For instance, thicker leaves with smoother textures not only enhance light absorption but also provide a stable biochemical environment for chlorophyll accumulation (Chang *et al.*, 2018). Similarly, rougher leaf surfaces correlate with higher flavonoid levels, reflecting their role in UV protection and oxidative stress mitigation (Zivcak *et al.*, 2017).

The integration of GLCM features into ANN models provides a powerful tool for non-invasive, scalable, and data-driven plant biochemical assessment. Such an approach can facilitate sustainable agricultural practices, ensure consistent medicinal plant quality, and ultimately advance our understanding of how leaf structure shapes biochemical dynamics. For the pharmaceutical industry, the ability to standardize chlorophyll and flavonoid content in medicinal plants like *Vernonia amygdalina* ensures product consistency and quality, addressing a critical challenge in plant-based medicine production (Paponov *et al.*, 2021).

Conclusion: This study demonstrates that integrating Gray Level Co-occurrence Matrix (GLCM) texture features with Artificial Neural Networks (ANNs) provides a robust, non-

invasive approach to estimating chlorophyll and flavonoid content in *Vernonia amygdalina* leaves. By linking texture characteristics such as contrast, correlation, energy, and homogeneity to underlying biochemical properties, the GLCM-ANN framework effectively captures the complex, non-linear relationships between leaf structure and compound levels. Chlorophyll content peaked in older leaves and aligned with smoother textures, while flavonoid concentrations were greatest in leaves exhibiting more complex textural patterns. The optimized ANN structure, featuring two hidden layers with 30 nodes each, achieved high predictive accuracy without destructive sampling. These findings highlight the potential of this method for enhancing real-time monitoring, optimizing resource management in precision agriculture, and improving quality control in the pharmaceutical industry. This research enriches our understanding of plant phenotyping, paving the way for more sustainable, data-driven approaches to plant health assessment and bioactive compound standardization.

CRedit author statement: RD designed the study, contributed the analysis, and drafted the manuscript; YH supervised the study, reviewed data processing and the manuscript; S and BDA supervised the study and reviewed the manuscript. All authors approved contributed to the final approval of the submitted version.

Conflicts of interest: The authors declare that they have no conflicts of interest.

Acknowledgement: The authors wish to thank Rut Juniar Nainggolan and Mitha Sa'diyah for their work in assisting with sample data collection in the laboratory.

Funding statement: This research funded by a research grant program (Hibah Professor) from Universitas Brawijaya, Indonesia.

Ethical statement: All procedures conducted in this study complied with the relevant national and institutional guidelines governing the collection and use of plant materials. This research did not involve human participants or vertebrate animals.

Availability of data and material: The data underpinning the conclusions of this study can be obtained from the corresponding author upon a reasonable request.

Informed consent: Not applicable.

Consent to participate: All authors participated in this research study.

Consent for publication: All authors have read and approved the final version of this manuscript and consent to its submission and publication in JGIAS.



SDG's addressed: Industry, Innovation, and Infrastructure, Zero Hunger, Responsible Consumption and Production, Good Health and Well-being.

Policy referred: Precision Agriculture Policy Initiatives, Digital Transformation in Agriculture (Smart Farming), Support for Medicinal Plant Research & Quality Control, AI and Machine Learning Integration Policy.

Publisher's note: All claims stated in this article are exclusively those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated/assessed in this article or claimed by its manufacturer is not guaranteed or endorsed by the publisher/editors.

REFERENCES

- Adebisi, M., R. Ogundokun and A. Abokhai. 2020. Machine learning-based predictive farmland optimization and crop monitoring system. *Scientifica* 2020:9428281.
- Annadurai, R. and R. Nelson. 2018. Effect of application of vermicompost, cowdung, neem cake and biofertilizer on growth and yield responses of cotton (*Gossypium hirsutum* L.). *International Journal of Current Research in Biosciences and Plant Biology* 5:70-76.
- Arnon, I. 1949. Copper enzymes in isolated chloroplasts: Polyphenoloxidase in *Beta vulgaris*. *Plant Physiology* 24:1-15.
- Baker, Z. and S. Sardari. 2021. Molecularly imprinted polymer (MIP) applications in natural product studies based on medicinal plant and secondary metabolite analysis. *Iranian Biomedical Journal* 25:68-77.
- Benos, L., A. Tagarakis, G. Dolias, R. Berruto, D. Kateris and D. Bochtis. 2021. Machine learning in agriculture: a comprehensive updated review. *Sensors* 21:3758.
- Butnariu, M. and I. Sarac. 2018. Interdisciplinary character of biochemistry and analytical biochemistry. *Biochemistry & Analytical Biochemistry* 3:1-5.
- Chang, X., Y. Lu, L. Zhang, J. Qiu, X. Guo, J. Pan and A.M. Abbasi. 2018. Impact of leaf development stages on polyphenolics profile and antioxidant activity in *Clausena lansium* (Lour.) Skeels. *Biomed Research International* 2018:1-8.
- Derbalah, A. 2012. Efficacy of some botanical extracts against *Trogoderma granarium* in wheat grains with toxicity evaluation. *The Scientific World Journal* 2012:1-9.
- Guo, Y., T. Wang, F. Fu and G. Wang. 2020. Temporospatial flavonoids metabolism variation in *Ginkgo biloba* leaves. *Frontiers in Genetics* 11:1-13.
- He, Y., L. Pan, T. Yang, W. Wang, C. Liu, B. Chen and Y. Shen. 2021. Metabolomic and confocal laser scanning microscopy (CLSM) analyses reveal the important function of flavonoids in *Amygdalus pedunculata* pall leaves with temporal changes. *Frontiers in Plant Science* 12:1-10.
- Kim, J. 2020. Minimizing diurnal variations in plant biochemical studies. *Journal of Plant Ecology* 17:134-148.
- Kumar, J., R. Joshi, R. Sathasivam, B. Nguyen, M. Faqeerzada, S. Park and B. Cho. 2022. Non-destructive measurement of total phenolic compounds in *Arabidopsis* under various stress conditions. *Frontiers in Plant Science* 13:20.
- Li, X., N. Park, H. Xu, S. Woo, C. Park and S. Park. 2010. Differential expression of flavonoid biosynthesis genes and accumulation of phenolic compounds in common buckwheat (*Fagopyrum esculentum*). *Journal of Agricultural and Food Chemistry* 58:12176-12181.
- Mahmoud, G. 2023. Responses of pea (*Pisum sativum* L.) to single and consortium bio-fertilizers in clay and newly reclaimed soils. *Plants* 12:3931.
- Nabwire, C., J. Mwangi and P. Ochieng. 2021. Generalizability challenges in machine learning models for diverse plant species. *Journal of Computational Botany* 14:200-215.
- Ogunbo, J. N., O.A. Alagbe, M.I. Oladapo and C. Shin. 2020. N-Hidden Layer Artificial Neural Network Architecture Computer Code: Geophysical Application Example. *Heliyon* 6:e04108.
- Paponov, M., M. Antonyan and I. Paponov. 2021. Decoupling of plant growth and accumulation of biologically active compounds in leaves, roots, and root exudates of *Hypericum perforatum* L. by the combination of jasmonate and far-red lighting. *Biomolecules* 11:1283.
- Rivero-Montejo, S., M. Vargas-Hernandez and I. Torres-Pacheco. 2021. Nanoparticles as novel elicitors to improve bioactive compounds in plants. *Agriculture* 11:134.
- Roitsch, T., L. Cabrera-Bosquet, A. Fournier, K. Ghamkhar, J. Berni, F. Pinto and E. Ober. 2019. Review: new sensors and data-driven approaches—a path to next generation phenomics. *Plant Science* 282:2-10.
- Rusadi, I. 2023. Chitosan extracted from cricket (*Gryllus bimaculatus*) induced bioactive compound and pigment content in *Centella asiatica*. *IOP Conference Series Earth and Environmental Science* 1271:012069.
- Sahin, S., Y. Özkan and A. Demirci. 2022. Leaf sampling techniques for accurate biochemical analysis. *Journal of Plant Sciences* 14:210-225.
- Sampaio, P., A. Almeida and C. Brites. 2021. Use of artificial neural network model for rice quality prediction based on grain physical parameters. *Foods* 10:3016.



- Sheng, X., H. Chen, H., J. Wang, Y. Zheng, Y. Li, Z. Jin and J. Li. 2021. Joint transcriptomic and metabolic analysis of flavonoids in *Cyclocarya paliurus* leaves. ACS Omega 6:9028-9038.
- Sinha, A. 2014. Biochemical analysis of *Psophocarpus tetragonolobus* L. (winged bean) and its role on restoration of degraded land of Raniganj and Barjora coalmine areas of West Bengal, India. Journal of Applied and Natural Science 6:792-796.
- Victoire, D. 2023. Leveraging artificial intelligence for enhancing agricultural productivity and sustainability. Quing International Journal of Innovative Research in Science and Engineering 2:141-156.
- Yan, Y., J. Ren, J. Tschannerl, H. Zhao, B. Harrison and F. Jack. 2021. Nondestructive phenolic compounds measurement and origin discrimination of peated barley malt using near-infrared hyperspectral imagery and machine learning. IEEE Transactions on Instrumentation and Measurement 70:1-15.
- Zhang, P. and C. Shen. 2019. Choice of the Number of Hidden Layers for Back Propagation Neural Network Driven by Stock Price Data and Application to Price Prediction. Journal of Physics: Conference Series 1302:022017.
- Zhang, R., S. Koh, M. Teo, R. Bi, S. Zhang, K. Dev and M. Olivo. 2022. Handheld multifunctional fluorescence imager for non-invasive plant phenotyping. Frontiers in Plant Science 13:822634.
- Zivcak, M., K. Brücková, O. Sytar, M. Brestič, K. Olšovská and S. Allakhverdiev. 2017. Lettuce flavonoids screening and phenotyping by chlorophyll fluorescence excitation ratio. Planta 245:1215-1229.

